



MANUAL PARA LA ELABORACIÓN DE TABLAS DE VOLUMEN DE PINO

PROYECTO “DESARROLLO DE TÉCNICAS PARA ELABORAR TABLAS O ECUACIONES DE PREDICCIÓN VOLUMÉTRICA PARA DOS ESPECIES DE PINO DEL SUR-OCCIDENTE DE MICHOACÁN (CONAFOR-2003-C03-10594)”



**EFRAÍN VELASCO BAUTISTA
SALVADOR MADRIGAL HUENDO
IGNACIO VÁZQUEZ COLLAZO
FRANCISCO MORENO SÁNCHEZ
ANTONIO GONZÁLEZ HERNÁNDEZ**

NOVIEMBRE DEL 2005

CONTENIDO

1. INTRODUCCIÓN.....	3
2. JUSTIFICACIÓN.....	3
3. OBJETIVO.....	3
4. ETAPAS EN LA ELABORACIÓN DE TABLAS DE VOLUMEN.....	3
4.1. DEFINICIÓN DE LA ESPECIE Y DEL ÁREA DE ESTUDIO.....	4
4.2. DEFINICIÓN DEL TIPO DE TABLA VOLUMEN.....	4
4.3. DETERMINACIÓN DEL NÚMERO DE ÁRBOLES A MEDIR.....	5
4.4. TOMA DE DATOS DE CAMPO.....	5
4.4.1. MEDICIÓN DE ÁRBOLES DERRIBADOS.....	5
4.4.2. MEDICIÓN DE ÁRBOLES CON TELE-RELASCOPIO.....	6
4.5. OBTENCIÓN DE VOLÚMENES INDIVIDUALES.....	8
4.6. EVALUACIÓN DE LOS MODELOS (ELECCIÓN DEL MODELO DE MEJOR AJUSTE).....	10
4.7. ELABORACIÓN DE LA TABLA DE VOLUMEN.....	10
5. ELEMENTOS BÁSICOS DE REGRESIÓN LINEAL.....	10
5.1. REGRESIÓN LINEAL SIMPLE.....	10
5.2. REGRESIÓN LINEAL MÚLTIPLE.....	16
5.3. COMPROBACIÓN DE LA ADECUACIÓN DEL MODELO.....	21
5.3.1. GRÁFICA DE PROBABILIDAD NORMAL.....	22
5.3.2. GRÁFICA DE RESIDUALES EN FUNCIÓN DE LOS VALORES AJUSTADOS.....	22
5.4. CRITERIOS PARA LA SELECCIÓN DEL MEJOR MODELO.....	23
6. ECUACIONES TRADICIONALES PARA LA ELABORACIÓN DE TABLAS DE VOLUMEN.....	24
6.1. ECUACIONES ARITMÉTICAS.....	24
6.2. ECUACIONES LOGARÍTMICAS.....	25
7. EJEMPLO DE APLICACIÓN (STATISTICAL ANALYSIS SYSTEM-SAS).....	27
8. REFERENCIAS.....	30

1. INTRODUCCIÓN

El principal parámetro de interés en un inventario forestal es el volumen maderable existente en la masa forestal objeto de estudio, el cual usualmente se estima empleando alguna estrategia de muestreo, en donde comúnmente las unidades de muestreo son parcelas de área fija y los árboles son las unidades de estudio. El volumen a nivel de unidad de muestreo es la suma de los volúmenes de los árboles que la integran. Por lo que el volumen de la unidad de estudio es el elemento básico en el proceso de estimación.

Dadas las dificultades prácticas de obtener directamente el volumen de los árboles en pie, éste se estima mediante expresiones matemáticas conocidas como **ecuaciones de volumen**, las cuales permiten predecirlo en función de dimensiones del árbol fáciles de medir, básicamente el diámetro normal y la altura total. Las **tablas de volumen** son arreglos numéricos en donde el volumen es expresado bajo una combinación de un diámetro y una altura. Estas tablas se generan al sustituir valores de diámetros y alturas en las ecuaciones de volumen.

2. JUSTIFICACIÓN

Durante la realización del primer inventario nacional forestal se construyeron tablas de volumen para diversos grupos de especies y para prácticamente todas las regiones del país, algunas de las cuales se siguen usando actualmente. Por otro lado, resulta interesante pensar que las ecuaciones a nivel de especie proporcionen estimaciones más precisas, de hecho la Ley General de Desarrollo Forestal Sustentable (Art. 73) y su Reglamento (Art. 37) establecen que los Programas de Manejo para el aprovechamiento de recursos forestales maderables, deberán contener un estudio dasométrico, en el cual las estimaciones de las existencias volumétricas serán por especie. Por tal motivo existe la necesidad de las dependencias operativas o prestadores de servicios técnicos de contar con estas herramientas dasonómicas (**tablas de volúmenes por especie**).

3. OBJETIVO

Proporcionar los elementos técnicos básicos para la elaboración de tablas de volumen de especies de pino.

4. ETAPAS EN LA ELABORACIÓN DE TABLAS DE VOLUMEN

Las etapas fundamentales para la elaboración de tablas de volumen son las siguientes:

- 1.- Definición de la especie y del área de estudio.
- 2.- Definición del tipo de tabla volumen.
- 3.- Determinación del número de árboles a medir.
- 4.- Toma de datos de campo.
- 5.- Obtención de los volúmenes individuales.

- 6.- Evaluación de los modelos (elección del modelo de mejor ajuste).
- 7.- Elaboración de la tabla de volumen.

4.1. DEFINICIÓN DE LA ESPECIE Y DEL ÁREA DE ESTUDIO

Durante 1961 a 1985, periodo en que se llevaron a cabo los trabajos del primer inventario nacional forestal, se construyeron tablas de volumen para diversos grupos de especies y para prácticamente todas las regiones del país, algunas de las cuales se siguen usando actualmente en los programas de manejo forestal, y en las estimaciones de volumen maderable que se requieren a nivel regional.

No obstante, la Ley General de Desarrollo Forestal Sustentable (Art. 73) y su Reglamento (Art. 37) establecen que los Programas de Manejo para el aprovechamiento de recursos forestales maderables, deberán contener un estudio dasométrico, en el cual las estimaciones de las existencias volumétricas serán por especie.

Por lo anterior, en los trabajos de planeación para la elaboración de tablas de volumen debe considerarse como punto inicial cual o cuales son las especies forestales que se evaluarán.

La definición del área de estudio es otro aspecto que también requiere de especial atención, puesto que la aplicación espacial de las ecuaciones de volumen que se generen estará restringida por el área geográfica en donde se colecten los datos de campo. Hoy día, con la ayuda de los Sistemas de Información Geográfica (SIG) y con los Sistemas de Posicionamiento Global (GPS), resulta relativamente fácil ubicar los puntos de muestreo en donde se obtiene la información de campo.

El uso de ecuaciones de volumen fuera del área geográfica en donde se generen, puede conducir a una sub o sobreestimación considerable del volumen real, ya que ejemplares de la misma especie biológicamente responden a las condiciones ambientales en donde se desarrollan.

4.2. DEFINICIÓN DEL TIPO DE TABLA DE VOLUMEN

Se pueden elaborar tablas de volumen fustal, o bien, tablas de volumen total.

Las tablas de volumen fustal son aquellas que consideran únicamente el fuste o tronco del árbol, las cuales pueden dividirse a su vez en tablas de volumen fustal con o sin corteza. La estimación del volumen sin corteza resulta sumamente importante, puesto que el volumen de la corteza representa en promedio entre un 10 y un 20% del volumen del árbol (Prodan *et. al.*, 1997).

Las tablas de volumen total son aquellas que además de considerar el volumen del fuste toman en cuenta el volumen de las ramas. Este tipo de tablas pueden ser importantes para la estimación de la biomasa aérea.

En la fase de planeación debe quedar claro el tipo de tablas de volumen que se requiere, en virtud de que el número de variables que se medirán en cada árbol, es función del tipo de tablas que se desee. Así por ejemplo, el grosor de la corteza indudablemente debe medirse en los árboles derribados, si el volumen sin corteza es de interés.

4.3. DETERMINACIÓN DEL NÚMERO DE ÁRBOLES A MEDIR

Como en otros estudios por muestreo, este aspecto más que a un principio teórico obedece a una situación práctica. Sin embargo, los siguientes criterios pueden conducir a la obtención de un número apropiado de árboles a medir:

1.- La muestra de árboles debe ser representativa de la población objeto de estudio. Si el bosque de interés está integrado por árboles jóvenes y maduros, la muestra deberá constituirse por elementos de los dos tipos. Las calidades de estación presentes en el bosque deben estar también representadas en la muestra.

2.- La muestra deberá distribuirse a través de la población objeto de estudio. Una muestra de árboles obtenida solamente en una zona aislada no reflejará las condiciones reales del área geográfica total de estudio, y por lo tanto su utilidad será dudosa.

3.- La muestra deberá incluir árboles de todas las categorías diamétricas. Las categorías diamétricas de mayor frecuencia existentes en el bosque deberán ser también las más frecuentes en la muestra. Se recomienda medir al menos diez árboles de las categorías diamétricas de mayor frecuencia.

4.- El número de árboles requerido para obtener una buena tabla de volumen aumentará con la variación del diámetro normal y de la altura.

5.- El número de árboles a medir deberá garantizar el cumplimiento de los supuestos básicos de la regresión.

6.- Por lo anterior, se recomienda medir al menos 215 árboles distribuidos en las diferentes categorías diamétricas y calidades de estación. Si el objetivo es estimar el volumen fustal sin corteza, es deseable que al menos el 35% de estos árboles se midan directamente.

4.4. TOMA DE DATOS DE CAMPO

4.4.1. MEDICIÓN DE ÁRBOLES DERRIBADOS

Esta técnica consiste en medir el diámetro normal (cm) estando el árbol en pie. Posteriormente el árbol se derriba y las trozas se miden directamente. Los diámetros de las trozas pueden medirse con forcípula y su longitud con un flexómetro o longímetro (Figura 1). En los aprovechamientos forestales la longitud

de las trozas ya está definida, por ejemplo, 8 ó 16 pies, sin embargo, siempre es deseable medir trozas pequeñas, ya que el cálculo del volumen será más exacto. Si el objetivo es estimar el volumen fustal sin corteza, el grosor de ésta debe medirse también en los extremos de las trozas.



Figura 1. Medición de trozas en árboles derribados.

Si bien es cierto que esta forma de obtener la información de campo conduce a resultados más exactos, tiene la desventaja de que demanda una gran cantidad de tiempo. Adicionalmente, los datos solamente pueden obtenerse en las áreas o predios sujetos a aprovechamiento forestal.

4.4.2. MEDICIÓN DE ÁRBOLES CON TELE-RELASCOPIO

El avance de la Dasimetría ha hecho posible la construcción de aparatos que permiten medir diámetros y longitudes en cualquier sección a lo largo del fuste de los árboles sin necesidad de apearlos. De estos aparatos, el Tele-relascopeo es sin lugar a dudas el más usado (Figura 2). Su utilidad se basa en que solo se requieren dos observaciones en cada extremo de las trozas simuladas. La primera observación se refiere al número de unidades taquimétricas en las cuales el diámetro de interés queda comprendido y la segunda a la pendiente de dicho diámetro con respecto a una línea horizontal imaginaria cuyo origen es el ojo del observador (Figura 3). La distancia del observador al árbol debe ser registrada.

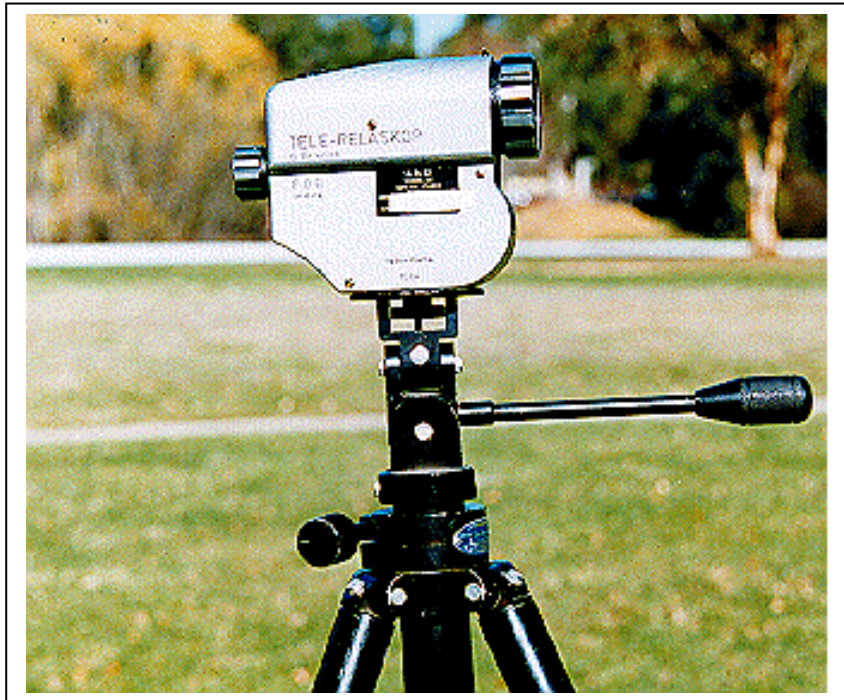
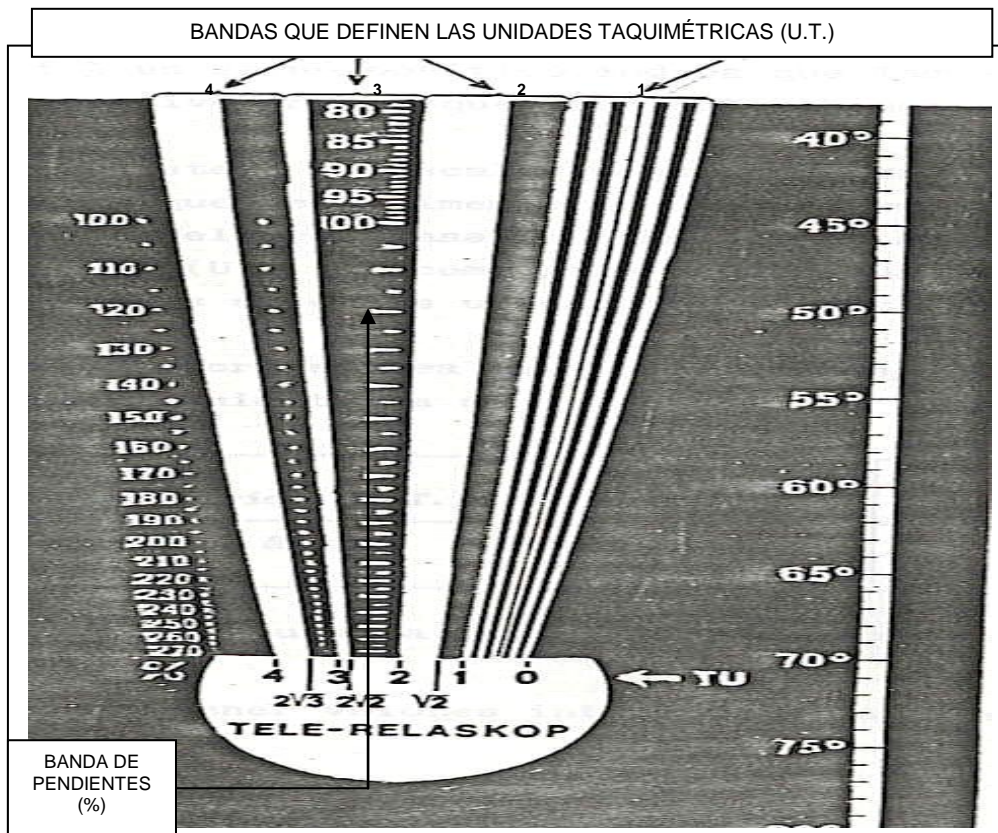


Figura 2. Tele-relascopio.



LA BANDA 1 ESTÁ DIVIDIDA EN DÉCIMOS. LAS BANDAS 2, 3 Y 4 ESTÁN INTEGRADAS POR UNA FRANJA BLANCA Y UNA NEGRA.

Figura 3. Bandas del Tele-relascopio.

La ventaja de este método radica en que la información se obtiene rápidamente, sin embargo, el grosor de la corteza no puede obtenerse en los extremos de las trozas simuladas. Además, los diámetros y las longitudes de las trozas, no se obtienen directamente.

Por lo anterior, se recomienda combinar ambos métodos, es decir, medir árboles directamente en aprovechamientos forestales y con Tele-relascopio. Es deseable que al menos el 35% de los árboles se midan con el primer método.

4.5. OBTENCIÓN DE VOLÚMENES INDIVIDUALES

Una vez que las trozas han sido medidas el siguiente paso es obtener su volumen (m^3). Todas las trozas medidas de manera directa pueden cubicarse mediante la fórmula de Smalian, cuya expresión matemática es la siguiente:

$$VT_i = \frac{L_i(S_{io} + S_{il})}{2}$$

Donde:

VT_i : Volumen de la i -ésima troza (m^3).

L_i : Longitud de la i -ésima troza (m).

$S_{io} = \frac{\pi D_{io}^2}{4}$: Área mayor de la i -ésima troza (m^2).

$S_{il} = \frac{\pi D_{il}^2}{4}$: Área menor de la i -ésima troza (m^2).

En consecuencia, el volumen fustal del árbol es la suma de los volúmenes de las trozas que lo integran, es decir:

$$VA = \sum_{i=1}^t VT_i$$

Donde:

t : Número de trozas del árbol.

De manera análoga, la altura del árbol es la suma de las longitudes de las trozas que lo integran.

Los diámetros de las trozas medidas con el Tele-relascopio se obtienen multiplicando las unidades taquimétricas que comprenden el diámetro de interés por la distancia a partir de la cual se realiza la lectura. Por ejemplo, si el diámetro observado queda comprendida entre 1.5 U.T. y el árbol se observa a una distancia de 21 m, el diámetro será de 31.5 cm (Figura 4).

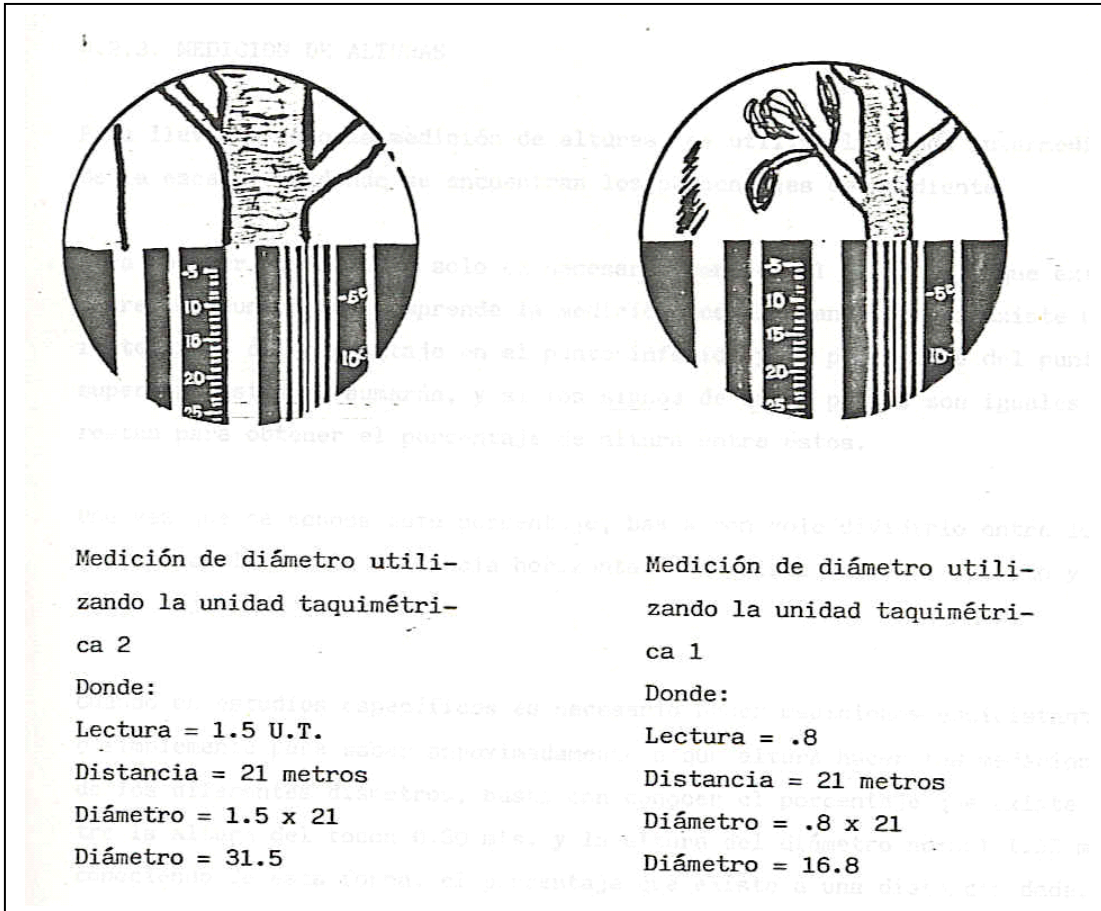


Figura 4. Obtención de diámetros a partir de Unidades Taquimétricas (U.T.).

La longitud de las trozas se obtiene como la diferencia entre las distancias multiplicadas por las pendientes cuando éstas son del mismo signo o como la suma cuando las pendientes son de signos diferentes.

Las trozas medidas con el Tele-relascopio también se cubican con la fórmula de Smalian, excepto la punta, cuyo volumen se obtiene mediante la fórmula del paraboloides apolónico, es decir,

$$VT_i = \frac{L_i S_{i0}}{2}$$

Con la información de campo se debe diseñar una base de datos en donde las columnas serán el número del árbol, el diámetro normal, la altura total y el volumen, mientras que los registros corresponderán a los diferentes árboles medidos.

4.6. EVALUACIÓN DE LOS MODELOS (ELECCIÓN DEL MODELO DE MEJOR AJUSTE).

En el ámbito forestal existen varias ecuaciones que pueden evaluarse para ver cual es la que mejor se ajusta a los datos en un caso particular. Los criterios para seleccionar el mejor modelo son de carácter netamente estadístico, por lo que serán presentados una vez que los principios de regresión sean explicados.

4.7. ELABORACIÓN DE LA TABLA DE VOLUMEN.

Una vez elegido el modelo que mejor se ajusta a los datos, en base a los coeficientes de regresión estimados se procede a calcular el volumen fustal del árbol, bajo la combinación de una categoría diamétrica y de altura dados. Esta actividad puede realizarse fácilmente empleando una hoja de cálculo o algún programa automatizado.

5. ELEMENTOS BÁSICOS DE REGRESIÓN LINEAL.

5.1. REGRESIÓN LINEAL SIMPLE.

El modelo de regresión lineal simple es un modelo con un solo regresor x que tiene una relación con una respuesta y , donde la relación es una línea recta. Este modelo es:

$$y = \beta_0 + \beta_1 x_1 + \varepsilon \quad (1)$$

Donde la ordenada al origen β_0 y la pendiente β_1 son constantes desconocidas, y es ε un componente aleatorio de error. Se supone que los errores tienen promedio cero y varianza σ^2 desconocida. Además, se suele suponer que los errores no están correlacionados.

El regresor x está controlado por el investigador, y se puede medir con error despreciable, mientras que respuesta y es una variable aleatoria. Con lo que hay una distribución de probabilidades de y para cada valor posible de x . La media de esta distribución es:

$$E(y / x) = \beta_0 + \beta_1 x \quad (2a)$$

y la varianza

$$Var(y / x) = Var(\beta_0 + \beta_1 x + \varepsilon) = \sigma^2 \quad (2b)$$

A los parámetros β_0 y β_1 se les llama coeficientes de regresión. La pendiente β_1 es el cambio de la media de la distribución de y producido por un cambio unitario en x . Si el intervalo de los datos incluye $x = 0$, entonces, la ordenada al origen, β_0 , es la media de la distribución de la respuesta y cuando $x = 0$. Si no incluye al cero, β_0 no tiene interpretación práctica.

Para estimar β_0 y β_1 se usa el método de **Mínimos Cuadrados**. Esto es, se estiman β_0 y β_1 tales que la suma de los cuadrados de las diferencias entre las observaciones y_i y la línea recta sea mínima. Según la ecuación (1), se puede escribir:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, 2, \dots, n \quad (3)$$

y

$$\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i)$$

Se considera que la ecuación (1) es un **Modelo de Regresión Poblacional**, mientras que la ecuación (3) **es un Modelo de Regresión Muestral**, escrito en términos de los n pares de datos: $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$.

Los estimadores por mínimos cuadrados de β_0 y β_1 , que designan por $\hat{\beta}_0$ y $\hat{\beta}_1$, deben satisfacer:

$$\left. \frac{\partial S}{\partial \beta_0} \right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

y

$$\left. \frac{\partial S}{\partial \beta_1} \right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0$$

Simplificando estas dos expresiones se obtiene:

$$n \hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$$

$$\hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i$$

Estas se llaman **Ecuaciones Normales de Mínimos Cuadrados**. Su solución es la siguiente:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (4)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i - \frac{\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n x_i\right)}{n}}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}} \quad (5)$$

En donde:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad \text{y} \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Por consiguiente, $\hat{\beta}_0$ y $\hat{\beta}_1$ en las ecuaciones (4) y (5) son los **estimadores por Mínimos Cuadrados** de la ordenada al origen y la pendiente, respectivamente. El modelo ajustado de regresión lineal simple, es entonces:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 \quad (6)$$

La ecuación (6) produce una estimación puntual de la media de y para una determinada x .

La diferencia entre el valor observado y_i y el valor ajustado correspondiente \hat{y}_i se llama **Residual**. Matemáticamente, el i -ésimo residual es:

$$e_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i), \quad i = 1, 2, \dots, n \quad (7)$$

Los residuales tienen un papel importante para investigar la adecuación del modelo ajustado de regresión, y para detectar diferencias respecto a las hipótesis básicas.

Se puede demostrar que los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ por mínimos cuadrados son estimadores insesgados de los parámetros β_0 y β_1 del modelo.

La varianza de $\hat{\beta}_1$ es:

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma^2}{S_{xx}} \quad (8)$$

Donde:

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$$

y la varianza de $\hat{\beta}_0$ es:

$$Var(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \quad (9)$$

Adicionalmente se tiene que un estimador insesgado de σ^2 es:

$$\hat{\sigma}^2 = \frac{SS_{Res}}{n-2} = MS_{Res} \quad (10)$$

Donde:

$$SS_{Res} = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

La cantidad MS_{Res} se llama **Cuadrado Medio Residual**.

Supóngase que se desea probar la siguiente hipótesis:

$$\begin{aligned} H_0 : \beta_1 &= 0 \\ H_1 : \beta_1 &\neq 0 \end{aligned} \quad (11)$$

Estas hipótesis se relacionan con la **significancia de la regresión**. El no rechazar $H_0 : \beta_1 = 0$ implica que no hay relación lineal entre x y y . Si se rechaza $H_0 : \beta_1 = 0$, eso implica que x tiene valor para explicar la variabilidad de y .

El procedimiento de prueba para $H_0 : \beta_1 = 0$ consiste en usar el estadístico t_0 , es decir:

$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} \quad (12)$$

Donde:

$$se(\hat{\beta}_1) = \sqrt{MS_{Res} / S_{xx}}$$

La hipótesis de significancia de la regresión se rechaza si $|t_0| > t_{\alpha/2, n-2}$ o si el valor de $p < 0.025$.

También se puede usar el método de **Análisis de Varianza** para probar el significado de la regresión. Para lo cual debe considerarse que:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (13)$$

The diagram illustrates the decomposition of the total sum of squares into regression and residual sums of squares. At the top, equation (13) is shown: $\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$. Three arrows point from three boxes below to the terms in the equation. The first box, labeled 'SUMA DE CUADRADOS TOTAL (SS_T)', points to the leftmost term. The second box, labeled 'SUMA DE CUADRADOS DE REGRESIÓN (SS_R) O DEL MODELO', points to the middle term. The third box, labeled 'SUMA DE CUADRADOS DE RESIDUALES (SS_{Res}) O SUMA DE CUADRADOS DEL ERROR', points to the rightmost term.

y que la suma de cuadrados de regresión se puede calcular también como:

$$SS_R = \hat{\beta}_1 S_{xy} \quad (14)$$

Donde:

$$S_{xy} = \sum_{i=1}^n y_i x_i - \frac{\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n x_i\right)}{n} = \sum_{i=1}^n y_i (x_i - \bar{x})$$

Para probar $H_0 : \beta_1 = 0$, se calcula el estadístico F_0 y se rechaza $H_0 : \beta_1 = 0$ si:

$$F_0 > F_{\alpha, 1, n-2} \text{ o si el valor de } p < 0.05.$$

En el Cuadro 1 se resumen el procedimiento indicado.

Cuadro 1. Análisis de varianza para probar el significado de la regresión.

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	F calculada
Regresión	$SS_R = \hat{\beta}_1 S_{xy}$	1	MS_R	$F_0 = \frac{MS_R}{MS_{Res}}$
Residual	$SS_{Res} = SST - \hat{\beta}_1 S_{xy}$	$n-2$	MS_{Res}	
Total	SS_T	$n-1$		

A la cantidad

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_{Res}}{SS_T}$$

se llama **Coefficiente de Determinación** y considera la proporción de la variación explicada por el regresor. Esta cantidad varía entre cero y uno, es decir, $0 \leq R^2 \leq 1$. Los valores cercanos a 1 implican que la mayor parte de la variabilidad de la variable respuesta está explicada por el modelo de regresión.

5.2. REGRESIÓN LINEAL MÚLTIPLE

En general, la variable de respuesta y se puede relacionar con k regresores x_1, x_2, \dots, x_k , o variables predictoras. El modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (15)$$

se llama modelo de **Regresión Lineal Múltiple** con k regresores. Los resultados son extensiones de los que se obtuvieron para la regresión lineal simple. Los parámetros $\beta_0, \beta_1, \dots, \beta_k$ se llaman **Coefficientes de Regresión**. El parámetro β_j representa el cambio esperado en la respuesta y por cambio unitario en x_j cuando todas las demás variables regresoras $x_i (i \neq j)$ se mantienen constantes. Por esta razón, a los parámetros β_j , se les llama coeficientes de regresión parcial.

En la mayor parte de los problemas del mundo real no se conocen los valores de los parámetros (los coeficientes β_j), ni la varianza del error σ^2 , y se deben estimar a partir de los datos muestrales. La ecuación o modelo de regresión ajustada se suele usar para pronosticar observaciones futuras de la respuesta y , o para estimar la respuesta media.

Se puede aplicar el método de mínimos cuadrados para estimar los coeficientes de regresión de la ecuación (15). Supongamos que se dispone de $n > k$ observaciones, y sea y_i la i -ésima respuesta observada, y x_{ij} la i -ésima observación del regresor x_j . Los datos aparecerán como en el Cuadro 2. Se supone que los errores ε del modelo tienen promedio cero, varianza σ^2 desconocida y que no están correlacionados.

Cuadro 2. Datos para la regresión lineal múltiple.

Observación	Respuesta	Regresores			
i	y	x_1	x_2	\dots	x_k
1	y_1	x_{11}	x_{12}	\dots	x_{1k}
2	y_2	x_{21}	x_{22}	\dots	x_{2k}
n	y_n	x_{n1}	x_{n2}	\dots	x_{nk}

Se supondrá que las variables regresoras x_1, x_2, \dots, x_k , son fijas, es decir, que son matemáticas o no aleatorias, y que se miden sin error.

Se puede escribir en la siguiente forma el modelo muestral de regresión que corresponde a la ecuación (15):

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i$$

$$= \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i, i = 1, 2, \dots, n \quad (16)$$

La función de mínimos cuadrados es:

$$S(\beta_0, \beta_1, \dots, \beta_k) = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij})^2 = \sum_{i=1}^n \varepsilon_i^2 \quad (17)$$

Al minimizar esta función respecto a $\beta_0, \beta_1, \dots, \beta_k$ y simplificar, se obtienen las siguientes **Ecuaciones Normales de Mínimos Cuadrados**

$$n \hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n x_{ik} = \sum_{i=1}^n y_i \quad (18)$$

$$\hat{\beta}_0 \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1} x_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n x_{i1} x_{ik} = \sum_{i=1}^n x_{i1} y_i$$

$$\begin{matrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{matrix}$$

$$\hat{\beta}_0 \sum_{i=1}^n x_{ik} + \hat{\beta}_1 \sum_{i=1}^n x_{ik} x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{ik} x_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n x_{ik}^2 = \sum_{i=1}^n x_{ik} y_i$$

Nótese que hay $p = k + 1$ ecuaciones normales, una para cada uno de los coeficientes desconocidos de regresión. La solución de las ecuaciones normales serán los estimadores por mínimos cuadrados $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$.

Es más cómodo manejar modelos de regresión múltiple cuando se expresan en notación matricial. Matricialmente el modelo expresado por la ecuación (16) es

$$y = X\beta + \varepsilon$$

Donde

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & & x_{2k} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

En general, y es un vector de $n \times 1$ de las observaciones, X es una matriz de $n \times p$ de los niveles de las variables regresoras, β es un vector $p \times 1$ de los coeficientes de regresión y ε es un vector $n \times 1$ de errores aleatorios.

Las ecuaciones normales de mínimos cuadrados en forma matricial son:

$$X'X \hat{\beta} = X'y \quad (19)$$

Para resolverlas se multiplican ambos lados de (19) por la inversa de $X'X$. Así el estimador de β por mínimos cuadrados es:

$$\hat{\beta} = (X'X)^{-1} X'y \quad (20)$$

siempre y cuando exista la inversa $(X'X)^{-1}$. La matriz $(X'X)^{-1}$ siempre existe si los regresores son **linealmente independientes**, esto es, si ninguna columna de la matriz X es combinación lineal de las demás columnas.

Es fácil de ver que la forma matricial de las ecuaciones normales (19) es idéntica a la forma escalar (18). Al escribir (19) con detalle se obtiene:

$$\begin{bmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \cdots & \sum_{i=1}^n x_{ik} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \cdots & \sum_{i=1}^n x_{i1}x_{ik} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \sum_{i=1}^n x_{ik} & \sum_{i=1}^n x_{ik}x_{i1} & \sum_{i=1}^n x_{ik}x_{i2} & \cdots & \sum_{i=1}^n x_{ik}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \cdot \\ \cdot \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1}y_i \\ \cdot \\ \cdot \\ \cdot \\ \sum_{i=1}^n x_{ik}y_i \end{bmatrix}$$

Si se hace la multiplicación indicada, se obtiene la forma escalar de las ecuaciones normales (18). En esta presentación se ve que $X'X$ es una matriz simétrica $p \times p$, y que $X'y$ es un vector columna de $p \times 1$.

El vector de valores ajustados \hat{y}_i que corresponden a los valores observados y_i es:

$$\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y = Hy \quad (21)$$

La diferencia entre el valor observado y_i y el valor ajustado correspondiente \hat{y}_i es el **Residual**, $e_i = y_i - \hat{y}_i$. Los n residuales se pueden escribir cómodamente con notación matricial como sigue:

$$e = y - \hat{y} \quad (22)$$

Se puede desarrollar un estimador de σ^2 a partir de la suma de cuadrados de residuales:

$$SS_{Res} = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = e'e$$

$$SS_{Res} = y'y - \hat{\beta}'X'y \quad (23)$$

La suma de cuadrados de residuales tiene $n - p$ grados de libertad, porque se estiman p parámetros en el modelo de regresión. El **Cuadrado Medio Residual** o **Cuadrado Medio de Residuales** es

$$\hat{\sigma}^2 = \frac{SS_{Res}}{n - p} = MS_{Res} \quad (24)$$

Este es un estimador insesgado de σ^2 .

Supóngase que se desea determinar si hay una relación lineal entre la respuesta y y cualquiera de las variables regresoras x_1, x_2, \dots, x_k . Este procedimiento suele considerarse como una prueba general o global. Las hipótesis pertinentes son:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0$$

$$H_1 : \beta_j \neq 0 \text{ al menos para una } j.$$

El rechazo de la hipótesis nula implica que al menos uno de los regresores x_1, x_2, \dots, x_k contribuye al modelo en forma significativa. El estadístico

$$F_0 = \frac{SS_R / k}{SS_{Res} / (n - k - 1)} = \frac{MS_R}{MS_{Res}}$$

Sigue una distribución $F_{k, n-k-1}$. Por consiguiente para probar $H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0$, se calcula el estadístico F_0 y se rechaza H_0 si:

$$F_0 > F_{\alpha, k, n-k-1} \text{ o si el valor de } p < 0.05$$

El procedimiento de prueba se resume en Cuadro 3.

Cuadro 3. Análisis de varianza para probar el significado de la regresión.

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	F calculada
Regresión	SS_R	k	MS_R	$F_0 = \frac{MS_R}{MS_{Res}}$
Residual	SS_{Res}	$n - k - 1$	MS_{Res}	
Total	SS_T	$n - 1$		

En donde:

La **Suma de Cuadrados de la Regresión** es:

$$SS_R = \hat{\beta}' X' y - \frac{\left(\sum_{i=1}^n y_i \right)^2}{n} \quad (25)$$

La **Suma de Cuadrados de Residuales o Suma Residual de Cuadrados** es:

$$SS_{Res} = y' y - \hat{\beta}' X' y \quad (26)$$

La **Suma de Cuadrados Total**:

$$SS_T = y' y - \frac{\left(\sum_{i=1}^n y_i \right)^2}{n} \quad (27)$$

Las hipótesis para probar la significancia de cualquier coeficiente individual de regresión, como por ejemplo β_j , son:

$$\begin{aligned} H_0 : \beta_j &= 0 \\ H_1 : \beta_j &\neq 0 \end{aligned} \quad (28)$$

Si no se rechaza $H_0 : \beta_j = 0$, quiere decir que se puede eliminar el regresor x_j del modelo. El estadístico de prueba para esta hipótesis es:

$$t_0 = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \quad (29)$$

Donde C_{jj} es el elemento diagonal de $(X'X)^{-1}$ que corresponde a β_j . Se rechaza la hipótesis nula $H_0 : \beta_j = 0$ si $|t_0| > t_{\alpha/2, n-k-1}$ o si el valor de $p < 0.025$.

5.3. COMPROBACIÓN DE LA ADECUACIÓN DEL MODELO

Las principales premisas que se han hecho hasta ahora al estudiar el análisis de regresión son las siguientes:

- La relación entre la respuesta y los regresores es lineal, al menos en forma aproximada.
- El término error tiene media cero.
- El término de error tiene varianza constante.
- Los errores no están correlacionados.
- Los errores tienen distribución normal.

En general, no se pueden detectar desviaciones respecto a las premisas básicas examinando los estadísticos básicos de resumen, como por ejemplo, t , F ó R^2 . Estas son propiedades globales del modelo, y como tal no aseguran la adecuación del mismo.

5.3.1. GRÁFICA DE PROBABILIDAD NORMAL

Un método sencillo de comprobar la suposición de normalidad es trazar una gráfica de probabilidad normal de los residuales. Es una gráfica diseñada para que al graficarse la distribución normal acumulada parezca una línea recta.

La Figura 5a muestra una gráfica de probabilidad normal “idealizada”. Observarse que los puntos caen aproximadamente sobre una recta. La parte b muestra una curva que va bruscamente hacia arriba y hacia abajo en los dos extremos, lo que indica que las colas de esta distribución son demasiado gruesas para poder considerarla como una normal. Al contrario, la parte c muestra un aplastamiento en los extremos, que es un comportamiento característico de las muestras tomadas de una distribución con colas más delgadas que la normal. Las parte d y e de la gráfica muestran patrones asociados con asimetría positiva y negativa, respectivamente.

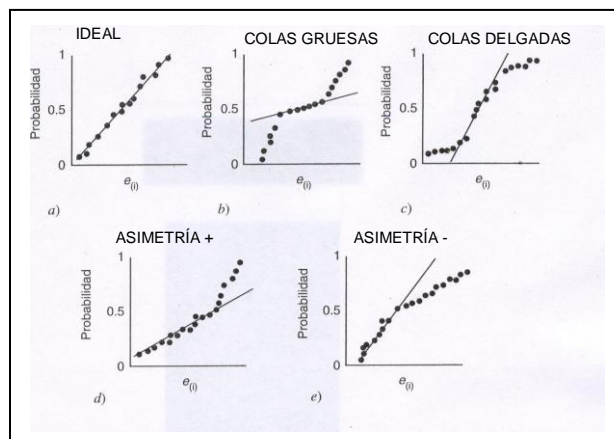


Figura 5. Gráficas de probabilidad normal.

Con frecuencia, los tamaños pequeños de muestra ($n \leq 16$) producen gráficas de distribución normal que se desvían bastante de la línea recta. Para muestras mayores ($n \geq 32$), las gráficas se comportan mucho mejor.

Otro criterio para verificar si existe normalidad de los errores es que el valor de p correspondiente a la prueba de Shapiro-Wilk sea mayor de 0.05.

5.3.2. GRÁFICA DE RESIDUALES EN FUNCIÓN DE LOS VALORES AJUSTADOS

Es útil una gráfica de los residuales e_i en función de los valores ajustados correspondientes \hat{y}_i . Si esta gráfica se parece como a la Figura 6a, que indica que los residuales se pueden encerrar en una banda horizontal, entonces no hay defectos obvios del modelo. Las distribuciones en las partes b y c indican que la varianza de los errores no es constante. La distribución en doble arco, como en la

parte c, se presenta con frecuencia cuando y es una proporción entre 0 y 1. El método común para manejar la no constancia de varianza es aplicar una transformación adecuada a la variable regresora o a la variable respuesta. Una gráfica en curva, como en la parte d, indica no linealidad. Esto podría indicar que se necesitan otras variables regresoras en el modelo. Por ejemplo, podría ser necesario un término al cuadrado. Las transformaciones de la variable regresora y/o la respuesta también podrían ayudar en estos casos.

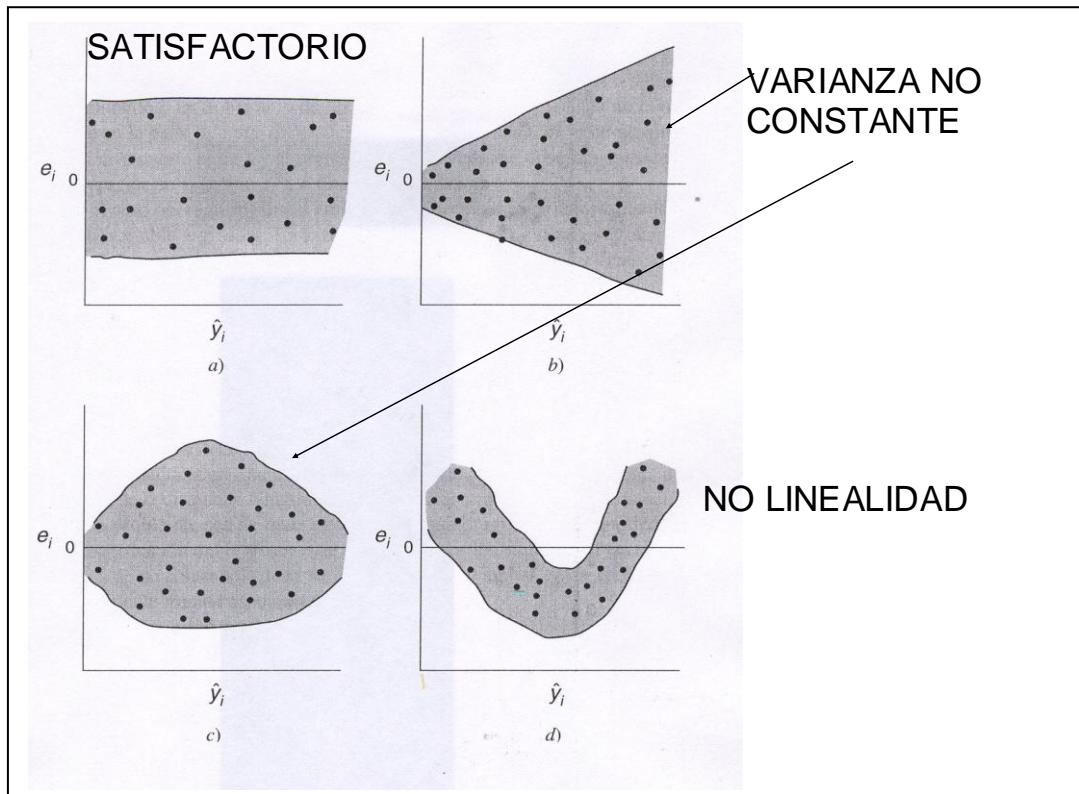


Figura 6. Patrones en las gráficas de residuales.

La no existencia de autocorrelación en los errores puede verificarse mediante el estadístico de Durbin-Watson. Valores relativamente cercanos 2 indican no autocorrelación (Pérez, 2003).

5.4. CRITERIOS PARA LA SELECCIÓN DEL MEJOR MODELO

Una vez que se ha verificado el cumplimiento de los supuestos básicos de regresión, y todos los modelos están en las mismas condiciones, para determinar el modelo que mejor se ajusta a los datos se consideran los siguientes criterios:

- a). Coeficiente de determinación R^2 (cercano a 1)
- b). Cuadrado medio del error (bajo)
- c). Coeficientes de regresión significativos (valor de $p < 0.05$ para la prueba de F y valor de $p < 0.025$ para la prueba de t).

d). Sencillez del modelo (menor número de variables independientes).

6. ECUACIONES TRADICIONALES PARA LA ELABORACIÓN DE TABLAS DE VOLUMEN

Por fortuna, como resultado de la experiencia de múltiples investigaciones en otros países, se ha llegado a concluir cuales son las ecuaciones más importantes que se ajustan en mayor grado a los datos de campo, mismas que se clasifican en aritméticas y logarítmicas (Caballero, 1972):

6.1. ECUACIONES ARITMÉTICAS

Son aquellas donde no intervienen logaritmos ni expresiones matemáticas complejas, como es la elevación de una variable a una potencia fraccionaria. Los exponentes que se utilizan en este tipo de expresiones son los dígitos uno y dos, aunque esto no excluye la utilización de otros números enteros. Es común separar aquellas ecuaciones aritméticas que emplean alguna medida de la forma de los árboles.

De acuerdo con lo anterior, las ecuaciones aritméticas más importantes se pueden resumir de la siguiente forma (Spurr, 1952; Husch, 1963):

Ecuaciones aritméticas que no consideran la forma del árbol

Del coeficiente mórfo constante

$$V = aD^2 A$$

De la variable combinada

$$V = a + bD^2 A$$

De la variable combinada modificada

$$V = a + bD^2 + cD^2 A$$

Australiana

$$V = a + bD^2 + cA + dD^2 A$$

Meyer modificada

$$V = a + bD + cDA + dD^2 + eD^2 A$$

Comprensible

$$V = a + bD + cDA + dD^2 + eA + fD^2A$$

De Naslund

$$V = a + bD^2 + cD^2A + dA^2 + eDA^2$$

De Tarata

$$V = \frac{D^2A}{a + bD}$$

Ecuaciones aritméticas que consideran la forma del árbol

Abreviada

$$V = a + bFD^2A$$

De la variable combinada

$$V = a + bF + cD^2A + dFD^2A$$

6.2. ECUACIONES LOGARÍTMICAS

Dentro de esta categoría se incluyen aquellas ecuaciones cuyo carácter exponencial permite expresarlas y resolverlas por medio del empleo de logaritmos.

También este tipo de ecuaciones se separan en dos grupos, según empleen o no alguna medida de la forma de árboles.

Las ecuaciones logarítmicas tradicionales son (Spurr, 1952; Husch, 1963):

Ecuaciones logarítmicas que no consideran la forma del árbol

De Schumacher

$$V = aD^b A^c$$

$$\log V = \log a + b \log D + c \log A$$

De Korsun

$$V = a(D+1)^b A^c$$

$$\log V = \log a + b \log (D+1) + c \log A$$

De Dwight

$$V = aD^b A^{3-b}$$

$$\log V = \log a + b \log D + (3-b) \log A$$

De la variable combinada

$$V = a(D^2 A)^b$$

$$\log V = \log a + b \log (D^2 A)$$

De Thornber

$$V = a(A/D)^b D^2 A$$

$$\log V = \log a + b \log (A/D) + \log (D^2 A)$$

Ecuaciones logarítmicas que consideran la forma del árbol

De forma a través del diámetro

$$V = aD^b A^c D_u^d$$

$$\log V = \log a + b \log D + c \log A + d \log D_u$$

De la variable combinada

$$V = a(FD^2 A)^b$$

$$\log V = \log a + b \log (FD^2 A)$$

En las ecuaciones anteriores:

V : Volumen del árbol

D : Diámetro normal

D_u : Diámetro del árbol en un punto más alto que el diámetro normal.

A : Altura del árbol.

F : Evaluación de la forma del árbol.

a, b, c, d, e y f son los coeficientes de regresión.

Las ecuaciones logarítmicas sin considerar la forma del árbol son generalmente las más usadas. De hecho, la transformación logarítmica conduce de manera natural a una distribución aleatoria de los residuales. La ecuación logarítmica de la variable combinada ha sido seleccionada por varios autores, incluyendo Hernández *et al.* (1999), Navarro *et al.* (2000), Cumplido (2002), Armendáriz *et al.* (2003), Madrigal y Vázquez (2004), Muñoz *et al.* (2005 y Velasco *et al.* (2005, inédito).

7. EJEMPLO DE APLICACIÓN (STATISTICAL ANALYSIS SYSTEM-SAS).

El establecimiento de un modelo de regresión es un proceso iterativo. Comúnmente se requiere de varios análisis a medida que se descubren mejoras en la estructura del modelo. Un buen programa de cómputo para regresión es una herramienta necesaria en este procedimiento. En este sentido se recomienda usar el Statistical Analysis System (SAS).

La estructura de un programa en SAS para estimar los parámetros de los modelos es el siguiente:

```
DATA JUNTAS;
INPUT ARBOL DIAMNORC ALTTOTAL VOLARBOL;
DIAMNORCM=DIAMNORC/100;
CARDS;
1      21.0   13.50  0.177
2      31.0   18.90  0.534
3      18.0   16.90  0.186
.
.
.
213    30.0   30.00  1.008
214    25.0   25.30  0.495
215    30.0   28.80  0.786

TITLE 'MODELO LOGARITMICO DE LA VARIABLE COMBINADA';
DATA LCOMBINA; SET JUNTAS; /*SE ANALIZA EL MODELO LOGARITMICO DE LA VARIABLE
COMBINADA*/;
LOGVOLARBOL=LOG(VOLARBOL);
DIAMNOR2=DIAMNORCM*DIAMNORCM;
DIAM2ALTU=DIAMNOR2*ALTTOTAL;
LOGDIAM2ALTU=LOG(DIAM2ALTU);
PROC PLOT; PLOT LOGVOLARBOL*LOGDIAM2ALTU='.';
PROC REG; MODEL LOGVOLARBOL=LOGDIAM2ALTU/DW P R; RUN;
```

Este programa hace lo siguiente:

```
DATA JUNTAS;
GENERA UN ARCHIVO QUE SE LLAMA JUNTAS.
INPUT ARBOL DIAMNORC ALTTOTAL VOLARBOL;
SE NOMBRAN LAS VARIABLES QUE SE ANALIZARÁN:
    DIAMNORC: DIAMETRO NORMAL (cm)
    ALTTOTAL: ALTURA TOTAL (m)
    VOLARBOL: VOLUMEN DEL ARBOL (m3)
```

DIAMNORCM=DIAMNORC/100; SE CREA UNA VARIABLE QUE SE LLAMA "DIAMNORCM", MEDIANTE LA CUAL SE OBTIENE EL DIÁMETRO NORMAL EN METROS.

CARDS;

SE INDICA QUE A CONTINUACION VENDRÁN LOS VALORES CORRESPONDIENTES A LAS VARIABLES.

SET JUNTAS;

SE INDICA QUE SE INVOQUE EL ARCHIVO JUNTAS.

LOGVOLARBOL=LOG(VOLARBOL);

SE GENERA UNA VARIABLE QUE SE LLAMA "LOGVOLARBOL" Y QUE ES IGUAL AL LOGARITMO NATURAL DE LA VARIABLE "VOLARBOL".

DIAMNOR2=DIAMNORCM*DIAMNORCM;

SE GENERA UNA VARIABLE QUE SE LLAMA "DIAMNOR2" Y QUE ES IGUAL AL "DIAMNORCM" (DIÁMETRO NORMAL EN METROS) AL CUADRADO.

DIAM2ALTU=DIAMNOR2*ALTTOTAL;

SE GENERA UNA VARIABLE QUE SE LLAMA "DIAM2ALTU" Y QUE ES IGUAL AL "DIAMNOR2" POR "ALTTOTAL".

LOGDIAM2ALTU=LOG(DIAM2ALTU);

SE GENERA UNA VARIABLE QUE SE LLAMA "LOGDIAM2ALTU" Y QUE ES IGUAL AL LOGARITMO NATURAL DE LA VARIABLE "DIAM2ALTU".

PROC PLOT; PLOT LOGVOLARBOL*LOGDIAM2ALTU='.';

SE INDICA QUE SE GRAFIQUEN LAS VARIABLES "LOGVOLARBOL" Y "LOGDIAM2ALTU".

PROC REG; MODEL LOGVOLARBOL=LOGDIAM2ALTU/DW P R; **RUN;**

SE OBTIENE LOS COEFICIENTES DE REGRESIÓN, EN DONDE "LOGVOLARBOL" ES LA VARIABLE DEPENDIENTE Y "LOGDIAM2ALTU" LA VARIABLE INDEPENDIENTE.

CON DW SE OBTIENE EL ESTADÍSTICO DE DURBIN-WATSON, CON P LOS VALORES PREDICHOS Y CON R LOS RESIDUALES.

RUN; INDICA QUE EL PROGRAMA SE EJECUTE.

Después de ejecutar un programa SAS, en donde se analizaron los modelos de Schumacher, Korsun, Logarítmico de la variable combinada y Thornber, con datos reales de *Pinus douglasiana*, se obtuvieron los resultados que se indican en el Cuadro 4.

En este Cuadro se puede observar que en general los modelos analizados para *P. douglasiana* cumplen con los supuestos clásicos de regresión, es decir:

- La varianza de los errores es constante u homocedástica (excepto en el modelo de Korsun), ya que la distribución de los residuales es aleatoria.
- No existe autocorrelación en los errores, ya que los valores del estadístico de Durbin-Watson están relativamente cercanos 2 (Pérez, 2003).
- Existe normalidad de los errores, puesto que en todos los casos el valor de p correspondiente a la prueba de Shapiro-Wilk es mayor de 0.05.

CUADRO 4. ESTADÍSTICOS DE RESUMEN DE LOS MODELOS EVALUADOS PARA ESTIMAR VOLUMEN FUSTAL EN FUNCIÓN DE DIÁMETROS INFERIORES O DE DIÁMETRO NORMAL Y ALTURA. ESPECIE *Pinus douglasiana*

MODELO	GL	CME	F _{cal}	PROB F	R ²	DR	DW	VALOR W	PROB W	$\hat{\beta}_i$	VALOR DE $\hat{\beta}_i$	EE DE $\hat{\beta}_i$	VALOR t	PROB t
SCHUMACHER	213	0.02237	7265.90	<.0001	0.9857	Aleatoria	1.555	0.993055	0.4137	$\hat{\beta}_0$	-0.33723	0.21175	-1.59	0.1128
										$\hat{\beta}_1$	2.08835	0.03818	54.70	<.0001
										$\hat{\beta}_2$	0.80587	0.05855	13.76	<.0001
KORSUN	213	0.04351	3683.72	<.0001	0.9722	Arco	1.478	0.992292	0.3232	$\hat{\beta}_0$	-5.59649	0.19183	-29.17	<.0001
										$\hat{\beta}_1$	6.94480	0.18329	37.89	<.0001
										$\hat{\beta}_2$	1.07858	0.07770	13.88	<.0001
LOGARITMICO DE LA VARIABLE COMBINADA	213	0.02332 (SCE=4.94440)	13927.7	<.0001	0.9850	Aleatoria	1.475	0.992292	0.3232	$\hat{\beta}_0$	-1.00655	0.01544	-65.19	<.0001
										$\hat{\beta}_1$	0.98957	0.00839	118.02	<.0001
THORNBUR	213	0.02237	7265.90	<.0001	0.9857	Aleatoria	1.555	0.993055	0.4137	$\hat{\beta}_0$	-0.33723	0.21175	-1.59	0.1128
										$\hat{\beta}_1$	-0.15887	0.05013	-3.17	0.0018
										$\hat{\beta}_2$	0.96474	0.01135	85.00	<.0001

DONDE:

GL: GRADOS DE LIBERTAD
 SCE: SUMA DE CUADRADOS DEL ERROR
 F_{cal}: VALOR DE "F" CALCULADA
 PROB F: PROBABILIDAD EXACTA DE "F"
 R²: COEFICIENTE DE DETERMINACIÓN
 DR: DISPERSIÓN DE RESIDUALES
 DW: ESTADÍSTICO DE Durbin-Watson

W: ESTADÍSTICO DE Shapiro-Wiks

\hat{B}_i : ESTIMADOR DE β_i
 FIV: FACTOR DE INFLACIÓN DE LA VARIANZA

EE: ERROR ESTÁNDAR DE \hat{B}_i
 PROB t: PROBABILIDAD EXACTA DE "t"

Aun cuando todos los modelos evaluados presentaron coeficientes de regresión significativos (valor de $p < 0.05$), valores de R^2 cercanos a 1 y CME bajos, por cierto muy similares, lo cual indica que cualquiera de éstos se ajusta bien a los datos, se seleccionó el modelo logarítmico de la variable combinada porque además de tener propiedades deseables estadísticamente, se trata de un modelo sencillo desde el punto de vista práctico.

Tomando en cuenta los coeficientes de regresión estimados, el modelo logarítmico de la variable combinada queda de la siguiente manera:

$$\text{LOG(VOLARBOL)} = -1.00655 + 0.98957 \text{ LOG(DIAM2ALTU)}$$

ó

$$V = 0.36548 (D^2 A)^{0.98957}$$

Donde:

V : Volumen fustal (m^3)
 D : Diámetro normal (m)
 A : Altura total (m)

8. REFERENCIAS

- Armendáriz O., R., A. Quiñónez Ch., M. Cano R., P. Juárez T., H. O. Rubio A. y J. Rentería A. 2003. Tablas de volúmenes para *Pinus herrerae* y *P. duranguensis* en el Ejido Monterde, municipio de Guazapares, Chih. Folleto Técnico No. 21. INIFAP. Chihuahua. 20 p.
- Caballero D., M. 1972. Tablas y tarifas de volúmenes: Análisis de algunas de las metodologías existentes y su aplicación a las condiciones de los bosques mexicanos. Nota I.N.F. No. 7. Subsecretaría Forestal y de la Fauna, SAG. México. 55 p.
- Cumplido O., R. 2002. Tablas de volúmenes y de incrementos para tres especies del género *Pinus* de tres predios del Estado de Chihuahua, México. Tesis de Licenciatura. Universidad Autónoma Chapingo. Chapingo, México. 114 p.
- Hernández M., E., P. Morales L., O. Delgado de J., E. H. Cornejo O., y S. Valencia Manzo. 1999. Predicción de volúmenes de fuste para *Pinus michoacana* Mart. y *Pinus douglasiana* Mart. en el Sureste de Nayarit. Foresta-AN. Nota Técnica No. 4. UAAAN. Saltillo, Coahuila. 9 p. <http://www.uaaan.mx/public/forestan/nota4.htm>. (11/07/2005).

- Huch, B. 1963. Forest mensuration and statistics. The Ronald Press Company. USA. 474 p.
- Madrigal H., S. e I. Vázquez C. 2004. Ecuaciones de predicción volumétrica para tres especies de *Pinus spp* en la Región Oriente de Michoacán. Ciencia Nicolaita. Revista Científica de la Universidad Michoacana de San Nicolás de Hidalgo. Vol. 39:189-206.
- Montgomery, D.C., E. A. Peck y G.G. Vining. 2004. Introducción al análisis de regression lineal. Trad. por V. González P. Compañía Editorial Continental. Primera reimpresión. México. 588 p.
- Muñoz F., H. J., S. Madrigal H., M. Aguilar R., J. García M. y M. Lara R. (2005). Tablas de volumen para *Pinus lawsonii* Roetzl. y *P. pringlei* Shaw. en el Oriente de Michoacán. Cien. For. en Méx. Vol. 28 (94):81-103.
- Navarro M., S. A., L. M. Torres E., A. Cano P., S. Valencia M. y E. H. Cornejo O. 2000. Predicción de volúmenes de fuste para *Pinus cembroides* Zucc., en el Sureste de Coahuila. Foresta-AN. Nota Técnica No. 3 UAAAN. Saltillo, Coahuila. 16 p. <http://www.uaaan.mx/public/forestan/nota3.htm>. (12/07/2005).
- Prodan, M., R. Peters, F. Cox y P. Real. 1997. Mensura forestal. Instituto Interamericano de Cooperación para la Agricultura (IICA). San José, Costa Rica. 561 p.
- Pérez L., C. 2003. Estadística. Problemas resueltos y aplicaciones. Pearson Prentice Hall. España. 485 p.
- Romahn de la V., C. F., H. Ramírez H. y J. L. Treviño G. 1994. Dendrometría. Universidad Autónoma Chapingo. Chapingo, México. 354 p..
- SEMARNAT. 2005. Ley general de desarrollo forestal sustentable y su reglamento. México. 267 p.
- Spurr, S. H. 1952. Forest inventory. The Ronald Press Company. USA. 476 p.
- Velasco B., E., S. Madrigal H., I. Vázquez C., F. Moreno S. y A. González H. Tablas de volumen con corteza para *Pinus douglasiana* y *P. pseudostrobus* del Sur-Occidente de Michoacán. Inédito. 17 p. (Artículo entregado a Ciencia Forestal en México el 14 de octubre del 2005).